

非文献コンテンツのための可視性と保守性に優れた 学術情報リポジトリの構築

著者	高田 良宏, 笠原 禎也, 西澤 滋人, 森 雅秀, 内島 秀樹
雑誌名	情報知識学会誌
巻	19
号	3
ページ	251-263
発行年	2009-10-01
URL	http://hdl.handle.net/2297/19715

非文献コンテンツのための可視性と保守性に優れた 学術情報リポジトリの構築

Construction of an Academic Resource Repository Excellent in Visibility and Maintainability for non-Bibliographic Contents

高田良宏^{1*}, 笠原禎也¹, 西澤滋人², 森雅秀³, 内島秀樹⁴

Yoshihiro TAKATA, Yoshiya KASAHARA, Shigeto NISHIZAWA,
Masahide MORI and Hideki UCHIJIMA

1* 金沢大学総合メディア基盤センター

Information Media Center, Kanazawa University

E-mail: yoshihiro@kenroku.kanazawa-u.ac.jp, kasahara@is.t.kanazawa-u.ac.jp

2 金沢大学自然科学研究科(現在, ソラン 株式会社)

Kanazawa University Graduate School of Natural Science & Technology(Currently, SORUN CORPORATION)

3 金沢大学人間社会学域人文学類

School of Humanities, College of Human and Social Sciences, Kanazawa University

E-mail: mori@kenroku.kanazawa-u.ac.jp

4 金沢大学情報部情報企画課

Information Infrastructure Service Division, Kanazawa University

E-mail: uchijima@ad.kanazawa-u.ac.jp

現在, 各地で構築・運用されている機関リポジトリは, 学術論文, 紀要などの文献のみを対象としている場合が多い. 一方, 写真, 動画, 観測データをはじめとした文献以外の学術情報は, 教育・研究上非常に有用なものでありながら対象外とされている場合が多い. 本研究では, 文献以外の学術情報をリポジトリ化する場合の問題点を整理し, その解決のため, メタデータの互換性, 異種コンテンツの共存, 大規模コンテンツの管理, コンテンツに関する位置情報を用いた可視化などを考案した. そして, 既存リポジトリプラットフォームのDSpaceを拡張し, 可視性と保守性に優れた汎用性の高い学術情報リポジトリの構築を行った. 本稿では, 構築したシステムの概要とその応用について述べる.

Recently, many institutional repositories mainly manage bibliographic material (for example, journal papers and bulletins). On the other hand, academic resources other than bibliographic material such as photographs, animations, and observational data are beyond the scope of the registration in many cases, though they are very useful for education and research. We studied problems in constructing a digital repository for non-bibliographic material, and proposed a way of managing numbers of digital contents, visualizing these contents by means of geographic information and offering a cross-reference environment over various metadata. We realized a general-purpose academic resource repository excellent in visibility and maintainability by improving DSpace, which is one of the conventional repositories. In the present paper, we introduce outline of the developed repository and its application to several kinds of non-bibliographic material.

キーワード: 学術情報リポジトリ, 機関リポジトリ, 非文献コンテンツ, 可視性, 保守性

Academic resource repository, Institutional repository, non-Bibliographic contents, Visibility, Maintainability

1. はじめに

大学には、数多くの学術情報が蓄積されている。さらに、大学の研究室では日々研究が行われており多くの学術情報が生産されている。このような学術情報を低コストで協調性を維持しつつ学内外に公開することを目的として、電子的な形態で集中的に蓄積・管理し、機関リポジトリとして公開する動きがある。国内では国立情報学研究所（NII）が中心となり推進してきた[1]。本学でも金沢大学学術情報リポジトリ（KURA）を立ち上げ運用を行なっている[2][3]。現在、運用されている機関リポジトリは、主に学術論文、紀要などの文献を対象としている。一方、写真、動画、観測データをはじめとした多くの貴重な学術情報は、教育・研究上有用なものでありながら対象外とされている場合が多い。この理由としては、コンテンツの種類が多種多様なため、主に書誌情報を元に情報管理を行うリポジトリへの登録にうまく適合しないことが挙げられる。また、これらのコンテンツの多くは、文書情報を含まず、全文検索が利用できないため、付加されたメタデータに対する検索が中心となり、検索性が低いと言わざるを得ない。このため、コンテンツ登録者には適切なメタデータ付けが、利用者には適切な検索語句の入力が求められ、これもリポジトリへの登録を阻む原因となっている。一方、学際的な研究や一般利用者へのアウトリーチに利用するには、異なる分野の多種多様なデータを、横断的、視覚的に検索できることが重要である。このような問題を解決し、文献以外のコンテンツをリポジトリ化できれば、多様な学術情報が教育・研究に有効利用されると期待

できる。なお、本論文では、学術論文、紀要、研究報告などの文献系のコンテンツを文献コンテンツ、それ以外の写真、動画、観測データなどのコンテンツを非文献コンテンツと記述する。

国内の機関リポジトリにおける非文献コンテンツの公開状況をみると、各機関の図書館などが登録している古書や貴重書の写真が挙げられる。しかし、古書や貴重書の写真は全文検索こそ出来ないが、メタデータの取り扱いを含め運用上は文献コンテンツと同様であり、今回対象とする多様な非文献コンテンツとは性質が異なる。また、千葉大学の CUWiC (Chiba University Wisdom Collection)[4]による衛星画像の公開例もあるが、CUWiC 自身はデータを蓄積しないいわゆるポータルリポジトリであり、専用サーバに蓄積されたメタデータを刈り取り公開している。現状では今回対象とする多様な非文献コンテンツを汎用的に取り扱えるリポジトリは存在しないといえる。

本研究では、文献以外の非文献コンテンツのリポジトリ化の指針となるべく、KURA などの文献リポジトリとも連携可能な、可視性と保守性に優れた汎用性の高い学術情報リポジトリの構築を目指した。まず、多種多様なコンテンツに汎用的に利用できるメタ情報の定義法を考案する。次に、多種多様で大量のコンテンツを体系的、かつ一括して登録する機能を考案する。さらにコンテンツの多くが地理的位置情報を有することに着目し、既存リポジトリプラットフォームに付加した地理的位置情報を介した検索機能を紹介する。

本論文の構成は次の通りである。2 章では非文献コンテンツのリポジトリ化におけ

る課題を整理し、開発条件を示す。続いて 3 章では構築したシステムの概要と課題の解決方法、4 章では他のリポジトリとの連携について述べる。5 章で応用と評価について述べ、最後に 6 章でまとめる。

2. 非文献コンテンツのリポジトリ化

2.1. 多様な非文献コンテンツ

1 章でも述べたが、大学内に蓄積している非文献コンテンツの多くは、その所有者や機関リポジトリを運用する図書館がリポジトリ化を強く望んでいるにも関わらず、リポジトリ化されていない。著者らが所属する金沢大学も例外ではなく、多様な非文献コンテンツを蓄積しており、それらのリポジトリ化が望まれているが実現されていない。表 1 に金沢大学内に蓄積され、本研究で考案するリポジトリの収録対象となり得る非文献コンテンツの一部について概要を示す。これらのコンテンツは各分野に広く存在し、件数は数千から数万に達するものも多い。種類は写真、動画、音声、バイナリデータなど多様である。また、非文献コンテンツの特徴として、全文検索が不可能なものが多いこと、コンテンツが生成された、あるいは、関連が深い地名などの情

報を持つものが多いことが挙げられる。次節では、表 1 に示したような非文献コンテンツをリポジトリ化する場合の課題を整理する。

2.2. リポジトリ化における課題

課題 1：機関リポジトリでコンテンツのメタデータ定義のために使用されている Dublin Core[5]（以下 DC とする）は、主に WWW 上でのリソースに関する情報を記述し、有用な情報の探索・発見に役立てる目的で制定されたものである。先行する文献リポジトリに関しては、DC の考え方に沿った記述法が確立されつつあるが、非文献コンテンツに対しては、それらがもつ多様で専門的な情報をどのように DC で定義するかが不明確である。このため、画像、動画、観測データといった各々のコンテンツを識別するために必要なメタデータが異なる非文献コンテンツに対して、メタデータをどのように互換性を維持しつつ定義するかを検討する。

課題 2：保守・管理面から考えた場合、管理者が保有する複数の性質の異なったコンテンツ（異種コンテンツ）をどのように管理するか、数千件、数万件以上におよぶコンテンツをどのように分類、登録・保守す

表1 蓄積されている非文献コンテンツ（抜粋）

コンテンツの内容	分野	種類	全文検索	件数等	地名等の情報
インドの仏像・壁画・遺跡	人文科学	写真	不可	2万件以上	発掘地/所蔵地
中国語の方言	人文科学	音声/地図	不可	約千件	調査地
中国の伝統芸能	人文科学	動画/写真	不可	800GB以上	撮影地
岩石標本	自然科学	写真	不可	数万件	採取地
「あけぼの衛星」の観測データ	自然科学	バイナリ/画像	不可	約7千件	観測座標
資料館所蔵品	共通	写真	不可	数千件	作成地/他
e-Learning素材	共通	作成ソフトに依存	一部可	約1万件	作成地

るか、などの問題がある。さらに、コンテンツ管理者が必ずしも情報技術の専門家であると限らないという問題もある。このため、情報技術に関して専門外である管理者でも、リポジトリの保守管理ができる仕組みの導入を検討する。

課題 3：前章で述べたが、非文献コンテンツは、文献コンテンツに比べ利用者に対する検索性が低いという問題がある。一方、非文献コンテンツには、発掘地、所蔵地をはじめとした地名などの情報を持つものが多い。地名などの情報は潜在的な地理的位置情報であり、地理的位置情報に変換すれば、地図上での視覚的な表現が可能となる。そこで、利用者に対する検索性を高めるため、地名などから地理的位置情報を得て、より視覚的に検索できる仕組みの導入を検討する。

課題 4：各リポジトリに存在する異種コンテンツ同士を横断的に検索、利用することを想定し、リポジトリ間の連携を行い、これらの情報をどのように相互参照するかという検討が必要である。リポジトリ間の連携は、OAI-PMH[6]プロトコルを用いたハーベスティングにより行われる。ハーベスティングとは、ネットワークを介して自動的に対象となるリポジトリのメタデータを刈り取る仕組みである。先行する文献リポジトリでは、ハーベスティングを行う場合のメタデータの扱い方など、その方法が確立されつつあるが、非文献コンテンツに対しても適用できるように検討する。

2.3. 開発方針

前節で挙げた課題 1～課題 4 を解決し、非文献コンテンツに対応した汎用性の高い学術情報リポジトリを構築するため、表 2

表2 開発条件表

汎用性の確保

①メタデータの互換性が確保できること
⇒ 当該リポジトリ上での詳細な定義と、他リポジトリとの互換性を両立 (課題1)

②他リポジトリとの連携を行えること
⇒ OAI-PMHプロトコルを活用した他リポジトリとの連携を実現 (課題4)

保守性の確保

③複数の異種コンテンツの管理を容易に行えること
⇒ 同一リポジトリ (同一システム) に異種コンテンツを容易に共存させる環境を導入 (課題2)

④多様かつ膨大な数のコンテンツの管理を容易に行えること
⇒ 分類の登録・管理機能、一括登録機能を導入 (課題2)

⑤プラットフォームの保守性が高いこと
⇒ 既存リポジトリプラットフォームをベースとする

可視性の向上

⑥コンテンツが持つ発掘地、所蔵地などの地理的位置に関する情報を可視化し、視覚的な検索機能を提供すること
⇒ 地理的位置情報とGoogle Earth[7]を連携させた検索機能を導入 (課題3)

に示す開発条件を設定した。

また、表 2 には、前節の課題には挙げていないが、⑤プラットフォームの保守性が高いこと、という項目を加え、既存リポジトリプラットフォームをベースにするという条件を設定した。これは、実運用を考えた場合、図書館などにおける文献コンテンツでの利用実績があり、保守体制などが確立しつつある既存リポジトリプラットフォームを利用することが、保守・管理コストの面で有利と判断したためである。そして、図 1 に示すように NII の学術機関リポジトリ構築連携支援事業の Web サイト[1]にリ

ンクされている機関におけるリポジトリプラットフォームの利用状況を調査したところ、DSpace[8]と DSpace をベースにしたプラットフォームが 77%を占めた。そこで、本研究では、当大学の KURA をはじめ国内で最も利用されている DSpace をベースに、機能を改良、追加する形で開発を進めることとした。

2.4. 使用データ

今回は、表 1 に挙げた非文献コンテンツの内、論文共著者の森がアジア図像集成[9]として蓄積しているインドの仏像・壁画・遺跡に関する画像を対象にして開発および検討を行った。図 2 左に、蓄積されているコンテンツの例を示す。アジア図像集成は、インドの各地で撮影された画像で、総数は 2 万件以上に及ぶ。従来、画像のメタデータは、データ管理者の森が、地域、タイトル、所蔵、特徴、サイズ、材質、制作年代などを Excel の表形式で管理していた。コンテンツの公開は、Web ページで行っており、手作業による作成・保守を行っていた。Web ページは地域により階層的に作成されており、地域からの検索はできるが、それ以外のメタデータを用いた検索機能は実装されていない。また、本システム開発後の検証の目的で、表 1 内の「あけぼの衛星」[10]の観測データ（図 2 中）および e-Learning 素材[11]（図 2 右）に適用することとした。

3. 学術情報リポジトリの構築

本システムは、リポジトリプラットフォームの DSpace をベースとして、機能を改良、追加する形で開発を進めた。開発にあ

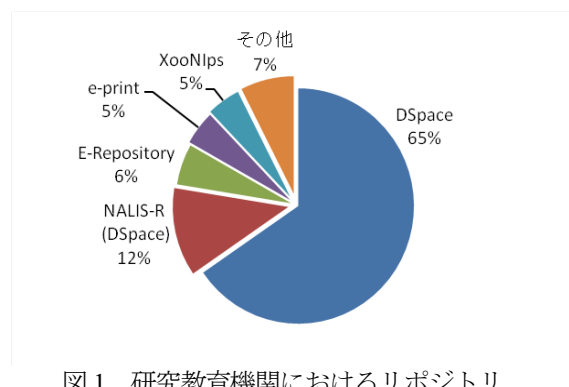


図1 研究教育機関におけるリポジトリプラットフォームの利用割合

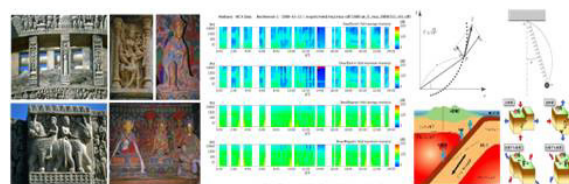


図2 蓄積されているコンテンツの例（左：アジア図像集成，中：あけぼの衛星の観測データ，右：e-Learning 素材）

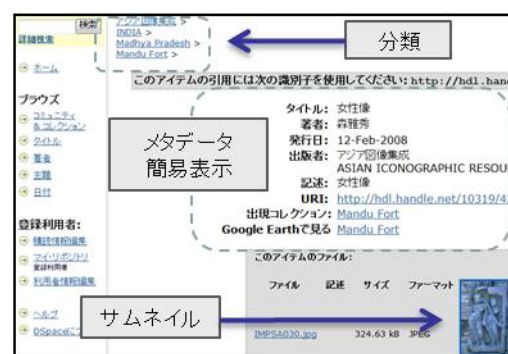


図3 アジア図像集成

たっては、DSpace の既存クラスを極力書き換えず、メソッドのオーバーライドおよび外部スクリプトを追加することとし、システムの移植や DSpace のバージョンアップといった場合にも極力影響が出ないように配慮した。図 3 は、実装したアジア図像集成の一面（情報表示画面）である[12]。本章では、メタデータの互換性の確保、保守性の確保および可視性の向上（表の 2 の課題 1～課題 3）の実現方法について述べる。

3.1. メタデータの互換性の確保

ハーベスティングによるリポジトリ間連携を考えた場合、メタデータ項目の追加を控え、DC の標準メタデータ語彙を用いることが望ましい。一方、教育・研究でのリポジトリ利用を考えると、各非文献コンテンツが持つメタ情報を詳細に記録することは必須である。しかし、DC の標準メタデータ語彙では、非文献コンテンツの情報を的確に表現することは困難なため、各々のコンテンツに合わせ、メタデータ項目を追加した拡張メタデータ語彙を定義する必要がある。

今回、各非文献コンテンツ特有のメタデータ項目の定義に Dumb-Down 原則[13]を導入した。Dumb-Down 原則とは、限定子を定義する場合、限定子を含めて記述したメタデータから限定子を取り除いてもメタデータと基本要素の間に矛盾が生じてはならないという規則である。Dumb-Down 原則は、組織間の運用ポリシーの違いなどによるメタデータ項目の差異を吸収するために用いられるが、ここでは、各非文献コンテンツの特性の違いによるメタデータ項目の差異を吸収するために用いた。

アジア画像集成特有の項目を表3に、定義した拡張メタデータ語彙を表4に示す。表4の網掛け部分が、新たに定義した項目である。後述するハーベスティングによるリポジトリ間連携を行った場合、拡張した限定子は取り除かれ、メタデータは基本要素に吸収され、基本要素として刈り取られる。拡張した基本要素はそれ自身が取り除かれメタデータは刈り取られない。例えば、表4中の材質は、次のように表される。

```
<description><material>砂岩
</material></description>
```

表3 アジア画像集成に必要なメタデータ項目（抜粋）

項目名	項目の説明	具体例
所蔵・所在	画像の所在地	東門南柱内側
出土地	画像の出土地	サンチー第一塔
材質	画像の材質	砂岩
サイズ	画像のサイズ	10m
撮影日	画像が撮影された年月日	2004/1/10
画像コード	管理者独自のコード体系	IMPSB050
ファイル名	画像のファイル名	2891-008
フォルダ名	画像の保管用フォルダ名	2891

表4 アジア画像集成用の拡張メタデータ語彙（抜粋）

基本要素	限定子	項目名
Coverage	currentLocation	所蔵・所在
	originalLocation	出土地
Description	material	材質
	dimensions	サイズ
Date	datePhotographed	撮影日
Identifier	original	画像コード
Appendix	nameOfFile	ファイル名
	nameOfFolder	フォルダ名

ハーベスティングを行なうと、拡張した限定子

```
<material></material>
```

が取り除かれ、

```
<description>砂岩</description>
```

となり、追加限定子が取り除かれても矛盾なく意味が通じることが保証される。また、表3中のフォルダ名やファイル名はコンテンツの管理のためだけに必要であり、ハーベスティングの際に刈り取られる必要はない、そのためこれらのメタデータは意図的

に基本要素として定義した。このように、Dumb-Down 原則に従ってのリポジトリ上での詳細な定義と、他リポジトリとの互換性を両立させることができる。

3.2. 保守性の確保

3.2.1. 異種コンテンツの共存

異種コンテンツを同一リポジトリ上に共存させるには、コンテンツごとに適したメタデータ語彙を設定できることと、コンテンツごとに適した表示を行えることが必須である。

本論文では、メタデータ語彙の設定や検索結果の一覧表示設定をコンテンツ種別に適切な設定に切り換えて利用できるようにコンテンツ別メタデータ語彙の設定法について考案した。図4に異種コンテンツを共存させた場合のイメージを示す。コンテンツの管理単位の最上位がルートコミュニティ（木構造最上部）であり、ルートコミュニティにはその下にサブコミュニティを何階層も作ることができる。アイテム（写真や動画などの実際のデータ）はコレクション（木構造最下部）に格納される。このような木構造を利用して、種類の異なるコンテンツごとにルートコミュニティを分け、それぞれのコンテンツのルートコミュニティごとにメタデータ語彙の設定と一覧表示の設定を行えるようにした。ルートコミュニティごとに設定を行う例として、一覧表示項目の設定法の概要を示す。DSpace では、一覧表示の表示項目の設定は、

定義名 = DC 語彙 1, DC 語彙 2, …

のように定義されている。DSpace に実装されている既存設定の設定方法を大幅に変更せず、簡単に設定が行えるよう、定義名にルートコ

ミュニティの識別番号を付加することで区別する仕組みとした。

定義名. ルートコミュニティの識別番号 =
DC 語彙 1, DC 語彙 2, …

異種コンテンツの共存を想定した環境は DSpace には実装されていないため、これらの仕組みを外部スクリプトおよび拡張ライブラリとして開発した。これにより同一リポジトリ上に各コンテンツの特性を反映した形での共存が可能となった。

3.2.2. コミュニティとコレクションの管理

前節で述べたが、通常のリポジトリシステムでは、コミュニティとコレクションで

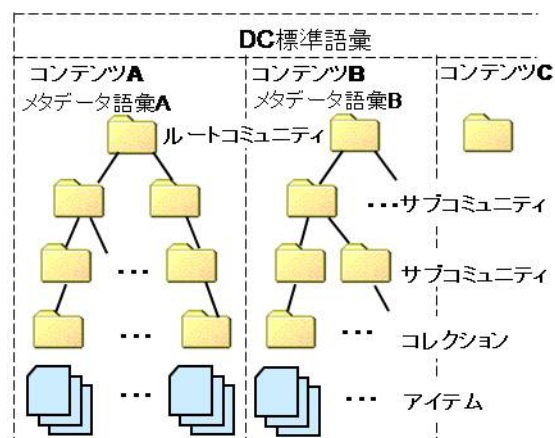


図4 異種コンテンツを共存させた場合のイメージ

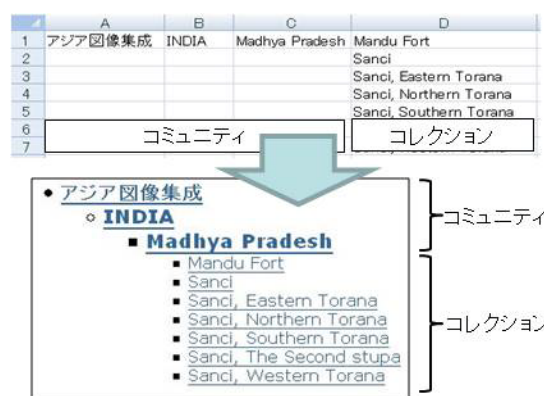


図5 コミュニティとコレクションの構造の記述例

アイテムを分類しコンテンツの管理を行っている。この管理のために Web インターフェイスが準備されるのが一般的であるが、非文献コンテンツの場合、コンテンツの分類が複雑なものが多く、Web インターフェイスを介した方法では体系的な管理が煩雑になりがちである。

これに対し本論文では、コンテンツ管理者（所有者）が情報技術の専門家でなくても比較的なじみやすい、Excel ファイルを用いたコミュニティとコレクション構造の管理法を考案した。図 5 は、アジア図像集成で用いた Excel ファイルの記述とリポジトリ上での表示例である。具体的には、図 5 に示すようなコミュニティとコレクションの構造を記述した Excel ファイルを読み込み、記述されている構造を解釈し、リポジトリ上の構造に反映させる。また、リポジトリ上のコミュニティとコレクションの構造を解釈し、Excel 形式で出力することも可能とした。この仕組みを DSpace に実装したことにより、コンテンツの分類体系が複雑な場合でも比較的容易に管理が可能となった。これにより、管理者のコミュニティとコレクション管理に掛かる負担が軽減されるとともに、リポジトリの構造の再現を容易に行うことが可能となった。

3.2.3. 一括登録

アイテムの登録においても、その数が数千件、数万件以上におよぶ非文献コンテンツに対して、Web インターフェイスで逐一登録する方法は、現実的とはいえない。

前節のコミュニティとコレクションの管理と同様に、コンテンツ管理者（所有者）が所有のコンテンツを管理するために、情報技術の専門家でなくても比較的なじみやすい Excel などの表計算ソフトを利用した

方法を考案した。表計算ソフト上で管理されるメタデータの出力（TAB 区切りまたは CSV 形式）を読み込み一括登録が行える仕組みである。具体的には、ファイルの内容を読み込み、メタデータなどの解析を行い、システムが理解できる形式に変更しリポジトリに登録を行うスクリプトを作成した。また、リポジトリに登録されているアイテムを TAB 区切りまたは CSV 形式で保存することも可能とした。図 6 に Excel などの表計算ソフト上でのメタデータの記述形式を示す。1 行目をヘッダ行として、メタデータ要素並びを記述し、2 行目以降に、各アイテムの情報を 1 行 1 アイテムとして、ヘッダ行に対応したメタデータ並び、登録先のコミュニティとコレクション名、アイテムが保存されているパスを入力する。この方法を DSpace の一括登録に導入することにより、管理者の登録の際の負担が軽減されるとともに、コンテンツごとの再現を容易に行うことが可能となった。

3.3. 位置情報を用いた情報の可視化

2 章でも述べたが、非文献コンテンツには、地名などの情報を持つものが少なくない。地名などの情報とは、地方・地域、市町村、山川湖、ランドマークなど（以降地名とする）が該当する。これらのコンテンツが持つ地名より地理的な位置情報を取得し、Google Earth と連携して、位置情報を地図上に表示することで可視化し、コンテ

表計算ソフトのワークシート					
1行目 ヘッダ行	メタデータ要素 1	メタデータ要素 2	...	メタデータ要素 n	コミュニティ・コレクション
2行目	1個目のアイテムの情報				アイテムへのパス
		
m+1行目	m個目のアイテムの情報				

図 6 メタデータの記述形式

ンツの検索性の改善を図った。今回、本システムと、Google Earth との連携には KML(Keyhole Markup Language)を用いた。KML は三次元地理情報を表現するための言語で、Google Earth や Google Map で地理情報を表すために用いられている。

アジア図像集成は、コレクション名が地名に相当するので、地名（＝コレクション名）から座標（経緯度）を取得することとして、2つの外部スクリプトを作成した。一つ目は、登録されている地名から座標を取得し、Google Earth 上に情報を表示させる KML を生成するスクリプトである。Google Earth 上に表示させる情報には、地名、説明文の他、アジア図像集成のコレクションへのリンクを含めることで、Google Earth 上からアジア図像集成への検索も可能とした。二つ目は、地図を表示するための視点を設定する KML を生成するスクリプトである。実装に外部スクリプトを用いたのは、DSpace への変更を最小限とするため、DSpace への変更は、アジア図像集成の情報表示画面内に Google Earth 上に情報を表示させるためのスクリプトへのリンクを埋め込んだことである。

地名から座標等の情報を得るために作成したテーブルの構造を表5に示す。定義した項目は、識別番号、地名、経度、緯度、標高、説明の6項目である。前述の通り、地名はコレクション名に対応しているので、コレクション単位で座標等の情報を返すこととなる。なお、標高と説明は、将来的な拡張を視野に入れて追加したもので、現在は未使用となっている。

Google Earth と連携した可視化システムの概要を図7に、また、Google Earth 上での表示例を図8に示す。Google Earth 上へ

表5 地理情報格納テーブルの構造

項目名 (列名)	項目の意味	具体例
geography_id	識別番号	6
location_name	地名	Sanci, The Second stupa
longitude	経度	77.44213
latitude	緯度	23.28445
altitude	標高	(現在未使用)
description	説明	(現在未使用)

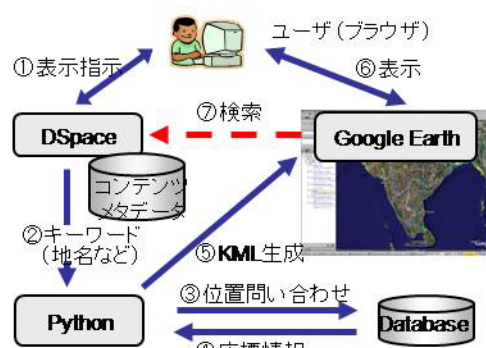


図7 Google Earth と連携した可視化の概要



図8 Google Earth 上での表示例

の表示およびGoogle Earth からDSpaceへの検索（リンク）は次のように行われる。なお、先頭の数字は図7中の矢印に付されている数字に相当する。

- ①DSpace に埋め込んだリンクをクリックし、Google Earth への表示を指示
- ②スクリプトに地名（＝コレクション名）

を渡す。これは、地図を表示した場合の視点の情報となる

- ③スクリプトは②の視点用の座標と全ての地名についての座標を問い合わせる
- ④視点用の座標と地名の座標を受け取る
- ⑤座標より地図上に情報を表示するための KML を作成し、それをユーザに返す
- ⑥Google Earth がインストールされているパソコンでは、Google Earth が起動し、地図上に情報を表示する
- ⑦地図上に表示される情報には DSpace のコレクションに関するリンクが埋め込んであり、Google Earth から DSpace への検索が可能である。

なお、Google Earth 上への表示に関しては、視点を地上から 5,000m の高さに設定し、選択された地名を中心に周辺を表示するモードとコンテンツに登録されている地名が全て範囲に収まるように表示するモードをリポジトリ（システム）ごとに選択することが可能である。

4. 他リポジトリとの連携

他リポジトリとの連携が可能であることを実証するために、学内の学術情報を統一的に公開するポータルリポジトリ（ハーベスタ）を構築した（表の 2 の課題 4）。これは、学内に立ち上げられたリポジトリ（プロバイダ）からメタデータをハーベスティングするハーベスタとしてリポジトリを構築し、個別のリポジトリにアクセスすることなく学内の学術情報を横断的に検索できるものである。図 9 にその概要を示す。プロバイダから収集されるのはメタデータの

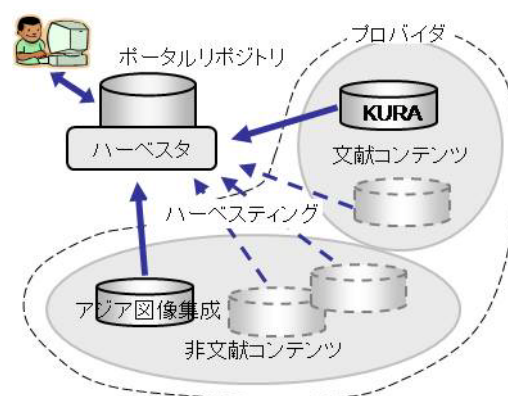


図 9 ハーベスティングによる統合検索

のみであり、アイテムは元のリポジトリのみに存在している。検索はポータルリポジトリで行なわれ、所望のアイテムが見つかった場合は、該当リポジトリへ移動する。なお、構築したポータルリポジトリは、ハーベスティングに OAI-PMH プロトコルを用いており、さらに、CNRI ハンドルシステム[14]に準拠しており、NII や他大学のリポジトリとも連携可能である。

今回は、非文献コンテンツに最適化した本システムと文献コンテンツに適した図書館の KURA で、非文献コンテンツと文献コンテンツを統一的に検索することができることを実証することができた。これにより、学術論文とそれに関係したデータ、図、写真、動画などの非文献コンテンツの一括検索や各リポジトリに存在する異なる分野のコンテンツの横断的な検索が可能となる。

5. 応用

本研究では、アジア図像集成の他に応用実装を行った。本章では、構築したリポジトリ単体での評価と運用における評価を行う。なお、応用実装として、あけぼの衛星の観測データおよび e-Learning 素材という特性の異なるコンテンツも同一の手法で構

築した。ただし、e-Learning 素材は学内で作成されたものであり、一大学のみでの運用では地名（作成地）が意味を持たないため Google Earth との連携を行っていない。図 10 は、あけぼの衛星の観測データの検索結果一覧画面（図左）と e-Learning 素材のアイテムの情報表示画面（図右）である。



図 10 応用実装（左：あけぼの衛星の観測データの一覧画面、右：e-Learning 素材の情報表示画面）

5.1. リポジトリ単体での評価

(1) メタデータ語彙

メタデータ語彙の定義法は、アジア画像集成とあけぼの衛星の観測データおよび e-Learning 素材に適用することで汎用性を確認することができ、非文献コンテンツへ適用する場合の一つの指針を示せたと言える。各分野で拡張メタデータ語彙を定義する場合でも、リポジトリ上での詳細な定義と他リポジトリとの互換性を両立させることができる。

(2) 保守性

コミュニティとコレクションの管理プログラムや一括登録スクリプトは、アジア画像集成とあけぼの衛星の観測データおよび e-Learning 素材という特性の異なるコンテンツに適用することで一般性と省力性を確認できた。また、異種コンテンツの共存は、単体で立ち上げたアジア画像集成とは別に、アジア画像集成とあけぼの衛星の観測データを共存させたリポジトリに適用することで本手法の有効性を確認できた。コミュニティとコレクションの管理プログラムは、DSpace にコンテンツの特徴に合わせて必要な構造を記述するだけで汎用的に適用可能である。一括登録スクリプトも、ヘッダ行に対応したメタデータ並び、登録先のコミュニティとコレクション名、アイテムの格納パスを指定するだけで、各種非文献コ

ンテンツに汎用的に適用可能である。

(3) 可視性

アジア画像集成が持つ地名を利用し、位置情報を Google Earth の地図上に表示することで可視化した。さらに、地図上に可視化した位置情報から DSpace への検索を可能とした。これにより、検索時の可視性と結果表示時の可視性が向上した。今回は、地名とコレクションを対応させることによって、コレクションから位置情報を取り出し可視化を行っている。同一構造のコンテンツであれば地名に対応する地理的な位置情報を準備するだけで適用可能である。文化遺産などの所蔵地、岩石標本や動植物標本の採取地をはじめ多くの非文献コンテンツが地名を持つことがわかっており、これらの位置情報を持つコンテンツに適用可能であると考えられる。

5.2. 運用

アジア画像集成は 2008 年 2 月、e-Learning 素材は 2008 年 7 月、あけぼの衛星の観測データは 2008 年 11 月に運用を開始し、現在も運用を続けている。

コミュニティとコレクションおよびアイテムの登録・管理に、一般の利用者がなじみやすい表形式を採用したことで、管理者が管理のために特別な技術を習得する必要は

なかった。今回使用した非文献コンテンツは、従来からメタデータを表形式で管理していたため、コミュニティとコレクションおよびアイテムの登録作業は非常にスムーズであった。検索については、既存プラットフォームをベースとしたため、文献コンテンツも含めコンテンツの特性が異なっても、基本的に同じ操作で検索を行うことができ、情報システムに不慣れな人に負担とならず使い易いものとなった。この間の運用やデモンストレーションを実施した結果、本システムは十分実用的なものであり、個人の情報技術に対する知識や技能の差に関係なく十分利用可能であると判断できた。また、異種コンテンツの共存と Google Earth による可視化機能により、異種コンテンツを同一地図上に表示させる利用法は、分野を超えた複数種のコンテンツの相互比較・参照が可能で、教育・研究への新たな活用法を示すことができた。

5.3. 課題

(1) 異種コンテンツの共存

今回、同一リポジトリ上に、複数の異種コンテンツを容易に共存させることが可能となった。しかし、特性が大きく異なったコンテンツを共存させた場合、リポジトリの利用形態をどのように想定するかで設定が大きく異なってくる。専門性を重視すると、各コンテンツの特性を重視したコンテンツ間の独立性が高い構造となり、その分一般利用者の使い勝手が悪くなったり、横断的な検索がしにくいものとなる。一方、一般性を重視すると、メタデータ項目の一般化が行われ、コンテンツ間の融合性は高くなるが、専門家には物足りないものになりがちである。今後、いかにして専門性と

一般性のバランスを取っていくか、さらに、いかにして専門性と一般性を両立させることができるかが課題である。

(2) 可視性

現在、地名と位置情報（座標）の登録を手動で行っており、登録コストが高いという問題がある。地理情報システムと Web サービスを連携させ、位置情報を取得できる環境を構築するなど、登録コストをいかに抑えることができるかが課題である。

6. まとめ

本研究では、非文献コンテンツをより広く公開し、教育・研究に有効利用されることを目的に、既存プラットフォームの問題点を改善し、可視性と保守性に優れた汎用性の高い学術情報リポジトリの構築を行った。

まず、非文献コンテンツをリポジトリ化する場合の問題点を4つの課題として整理し、非文献コンテンツに対応した学術情報リポジトリの開発条件として示した。次に、現実の非文献コンテンツ（アジア図像集成）をリポジトリとして構築する過程および構築したリポジトリの運用において、課題の解決手法の考案とその有効性を確認した。そして、アジア図像集成とは特性の異なるあけぼの衛星の観測データおよび e-Learning 素材という2種類のコンテンツにも適用することで有効性と一般性を示した。

今後、この研究の成果を基に、多種多様な多くの非文献コンテンツがリポジトリ化され、異なる分野のデータを横断的、視覚的に検索できるようになれば、学際的な研究や一般利用者へのアウトリーチに利用さ

れるものと期待できる。

謝辞

本研究の一部は、科学研究費補助金（基盤研究 C，課題番号 20510006）のもとに行われた。

参考文献

- [1] 国立情報学研究所(NII): 「学術機関リポジトリ構築連携支援事業」, <http://www.nii.ac.jp/irp/> (2009 年 4 月 20 日参照)
- [2] 金沢大学: 「金沢大学学術情報リポジトリ(KURA)」, <http://dspace.lib.kanazawa-u.ac.jp/dspace/> (2009 年 4 月 20 日参照)
- [3] 橋 洋平: 「金沢大学学術情報リポジトリ KURA の構築と課題」, 大学図書館研究, Vol.79, pp.18-26(2007).
- [4] 千葉大学: 「CUWiC (Chiba University Wisdom Collection)」, <http://narihira.ll.chiba-u.jp/> (2009 年 4 月 20 日参照)
- [5] Dublin Core Metadata Initiative, <http://www.dublincore.org/> (2009 年 4 月 20 日参照)
- [6] Open Archives Initiative: “Open Archives Initiative Protocol for Metadata Harvesting”, <http://www.openarchives.org/pmh/> (2009 年 4 月 20 日参照)

- [7] Google: 「Google Earth」, <http://earth.google.com/> (2009 年 4 月 20 日参照)
- [8] DSpace Foundation: “DSpace.org”, <http://www.dspace.org/> (2009 年 4 月 20 日参照)
- [9] 金沢大学: 「アジア図像集成(Asian Iconographic Resources)」, <http://air.w3.kanazawa-u.ac.jp/>, html 版 (2009 年 4 月 20 日参照)
- [10] I. Kimura, et al: “VLF Observations by the Akebono (EXOS-D) satellite, J. Geomag. Geoelectr., Vol. 42, pp.459-478(1990).
- [11] 高田良宏, 笠原禎也, 佐藤正英, 鈴木恒雄, 松本豊司, 森祥寛: 「e-Learning 素材管理・再利用システムの開発」, コンピュータ&エデュケーション, Vol.20, pp.68-73(2006).
- [12] 金沢大学: 「アジア図像集成(Asian Iconographic Resources)」, <http://www.db02.db.kanazawa-u.ac.jp/dspace/>, DSpace 版 (2009 年 4 月 20 日参照)
- [13] 杉本 重雄: 「メタデータについて : Dublin Core を中心として」, 情報知識学会誌, Vol.10, No.3, pp.53-58(2000).
- [14] Corporation for National Research Initiatives,(CNRI): “HANDLE.NET - The Handle System”, <http://www.handle.net/> (2009 年 4 月 20 日参照)

(2009 年 6 月 30 日受付)
(2009 年 9 月 27 日採択)